MCG-ICT at MediaEval 2016: Verifying Tweets From Both Text and Visual Content

Juan Cao[1], Zhiwei Jin[1,2], Yazi Zhang[1,2], Yongdong Zhang[1]
[1]Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
{jinzhiwei, caojuan, zhangyazi, zhyd }@ict.ac.cn

ABSTRACT

The Verifying Multimedia Use Task aims to automatically detect manipulated and misleading use of web multimedia content. This year we have two important improvements: On the one hand, considering that the prediction based on a short tweet is unreliable, we propose a topic-level credibility prediction framework. Which exploits the internal relations of tweets belonging to same topic, and enhance the prediction precision by topic sampling and fusing the topic-level features. This method reaches the precision of 74%. On the other hand, by referencing the handbook of professional journalists to identify fake videos/images, we propose several discriminative multimedia features, and build a decision tree based on these pure visual features. The results show that it is effective for fake video detection. This method can get the precision of 69%. Finally, the result by fusing the text and visual model is 76%.

1. PROPOSED APPROACH

The paper presents the approach developed by MCG-ICT for the MediaEval 2016 Verification Multimedia Use task. The task intents to automatically predict whether the given Web multimedia content is real or fake by techniques. Compared with the data of 2015, this year has two important changes: Firstly, this year pay more attention to the small event than the breaking news. More than 59% events contain less than 10 tweets and 95% are less than 50 tweets. Compared with last year's 42.5 tweets per event, the small event verification is more challenging. We propose a topic-level verification framework, which can improve tweet-level results by fusing the topic-level features and topic-level predict result. Secondly, this year extended video content. Considering that there have many effective ways to manually verify the video content in News industry, so we try to transfer the human experiences into automatic techniques, and proposed a visual feature based decision tree. More details about the task can be found in [1].

1.1 Text Analysis Approach

Generally, a tweet is very short (no more than 140 words). So its meaning is incomplete, and the credibility prediction on message-level is unreliable. We propose a topic-level classification framework to reference the context tweets. Compared with an independent tweet, a topic can maintain principal information and also eliminate random noise. However, considering that the data of 2016 includes many small events, so we design a topic sampling step to further remove the noise of weak topics. The framework includes following four main parts:

Topics Clustering: In our data, each tweet contains a kind of visual content(videos or images). As defined in [1], A tweet's credibility is directly related to its visual content's credibility. So rather than performing clustering algorithms [2] [3], we directly take a video/image as a topic. Tweets containing the same video/image belong to the same topic.

Topic Layer Features: For each topic, we compute the average of its tweets' features as its features. Besides, we propose several statistic features which are listed in Table 1. Completed topic-level feature is the combination of the two kinds of features above.

Topics Sampling: A topic-level classifier is constructed based on the above topic-level features. To remove the noise brought by small topics, we sample topics with high confidences in the 10-fold cross validation process. The sampling should keep the balance between the fake and real topics.

Fusing Topic Layer Results: the topic classifier classifies each topic and gives a corresponding probability value indicating how likely it is to be fake. We add this probabil-

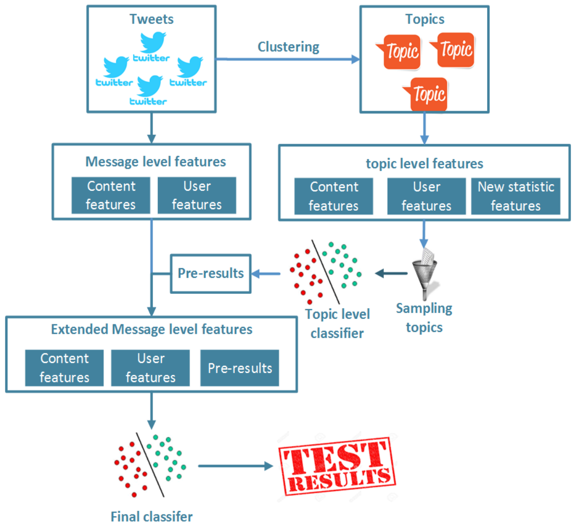

Figure 1: The framework of proposed two-level classification model. Topic level classification results are fused with message level to produce a final result.

MediaEval 2016 Workshop, Oct. 20-21, 2016, Hilversum, Netherlands.

Table 1: Topic Layer New Statistic Features

Feature	Explanation
num_tweets	the number of tweets
num_distinct_tweets/ hashtags	the number of distinct tweets /hashtags
distinct_tweets_index	the ratio of distinct tweets
contain_url/mention	the ratio of tweets containing urls/metions(namely, @)
contain_urls/mentions/ hashtags/questionmarks	the ratio of tweets containing multiple urls/mentions /hashtags/questionmarks

ity as another feature to the original features of its tweets. Finally, a message-level classifier is built on the extended features.

1.2 Visual Analysis Approach

We especially evaluate visual content's credibility for two reasons: Firstly, a tweet's credibility is directly defined by its visual content. Secondly, visual content is proved to be distinctive by journalists. So it's another significant feature besides text content.

In our dataset, Each tweet contains only one kind of visual content: images or videos. Observing that, we build two classification models on them separately: On the one hand, we propose 7 types of forensics feature [4] [5] [6] [7] to build the image model. On the other hand, as the main contribution of visual analysis, the video model is discussed in details for next three paragraphs.

In news industry, experienced journalists are usually precise in identifying videos. Referencing the handbook they write, we summarize four crucial features to detect fake videos: contrast ratio, clarity, time and a binary value indicating whether the video contains logos. To reproduce their process, we build a decision tree on these features which is illustrated in Figure 2. The framework including the following two main parts:

Logo Detecting: Logos are a kind of humanly-edited graphical symbols. They are often added on corners of videos to represent certain organizations. The basic idea to detect logos is that they are invariant compared with other parts of videos: we divide videos into frames, and detect color-fixed pixels in each frame. If a certain area's pixels keeps unchange for most frames, it is determined as a logo. To reduce random errors caused by steady dispersed pixels like short lines, we perform a median filter and only remain logos that pass the filter.

Clarity and Contrast Ratio Computing: A video's clarity and contrast ratio are two typical factors indicating its quality. For clarity, we use the traditional SMD algorithm [8] to compute it. For contrast ratio, we define it as the ratio of a video's size over its time.

2. RESULTS AND DISCUSSION

We submitted 3 results which are listed in Table 2. Run 1 only use text analysis approach while Run 2 only use visual analysis approach. Run 3 is a hybrid of text and visual approach: if a testing tweet contains videos/images, we use the visual model to classify it, otherwise we choose the text model.

From the results we can observe that:(1)Both two models reaches very promising results. (2) The text model is more effective than the visual model. We infer it's probably because of lacking sufficient videos and our video model is under-fitting. In the future, we want to explore more videos to validate and improve our model.

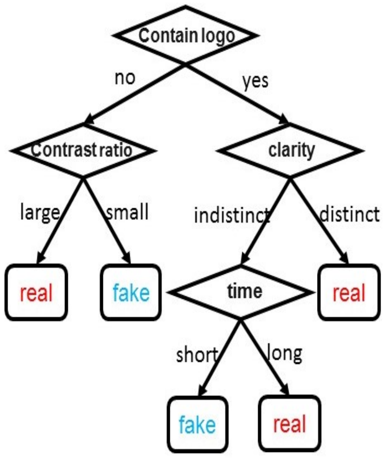

Figure 2: The video classification decision tree.

Table 2: Topic Layer New Statistic Features

	Recall	Precision	F1-Score
Run 1	0.629	0.747	0.683
Run 2	0.514	0.698	0.592
Run 3	0.610	0.764	0.678

3. ACKNOWLEDGMENTS

This work was supported by National Nature Science Foundation of China (61571424, 61172153) and the National High Technology Research and Development Program of China (2014AA015202).

4. REFERENCES

[1] Christina Boididou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, StuartE. Middleton, Andreas Petlund, and Yiannis Kompatsiaris. Verifying multimedia use at mediaeval 2016. In *Proceedings of the MediaEval 2016 Workshop*, Hilversum, Netherlands,Oct. 20-21, 2016.

[2] Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. News credibility evaluation on microblog with a hierarchical propagation model. In *2014 IEEE International Conference on Data Mining (ICDM)*, pages 230–239. IEEE, 2014.

[3] Tuan-Anh Hoang and Ee-Peng Lim. Microblogging content propagation modeling using topic-specific behavioral factors. In *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING(Volume: 28, Issue: 9, Sept. 1 2016)*, pages 2407 – 2422. IEEE, 2016.

[4] C.Pasquini, F.Perez-Gonzalez, and G. Boato. A benford-fourier jpeg compression detector. In *IEEE The International Conference on Image Processing, pp*, pages 5322–5326. IEEE, 2014.

[5] T.Bianchi and A.Piva. Image forgery localization via block-grained analysis of jpeg artifacts. In *IEEE Transactions on Information Forensics and Security, vol. 7, no. 3, 2012.*, pages 1003–1017. IEEE, 2012.

[6] M.Goljan, J.Fridrich, and M.Chen. Defending against fingerprint-copy attack in sensor-based camera identification. In *IEEE Transactions on Information Security and Forensics, vol. 6, no. 1, 2010.*, pages 227–236. IEEE, 2010.

[7] W.Li, Y.Yuan, and N.Yu. Passive detection of doctored jpeg image via block artifact grid extraction. In *ACM Signal Processing, vol. 89, no. 9, 2009.*, pages 1821–1829. IEEE, 2009.

[8] R.A. Javis. Focus optimization criteria for computer image processing. pages 163–180. Microscope, 1976.